

ТЕРМІН У СЕМАНТИЧНІЙ СТРУКТУРІ АНГЛОМОВНОГО НАУКОВОГО ТЕКСТУ

Скопюк Т.Г.

Київський національний університет імені Тараса Шевченка

Розглядається взаємозв'язок між системообумовленими семантичними властивостями наукового терміну (семантичною складністю та когнітивною вартістю) та його дискурсивними характеристиками, зокрема, приналежністю до різних компонентів лінійно-композиційної структури наукового тексту.

Система англійської термінології з обчислювальної техніки, як і будь-яка інша система, є семантично структурованою. Лексика субмови разом з її різноманітними семантичними зв'язками утворює структури різної складності та обсягу. У залежності від вибраних критеріїв у системі термінологічної лексики розрізняються певні структурні рівні: семантико-темантичний, стратифікаційний, тощо.

У попередніх публікаціях було зазначено, що існує різниця між когнітивною вартістю та інформативною місткістю тих термінів, що вживаються в різних структурних компонентах наукового тексту [Скопюк 1997: 40-45]. Цей факт є побічним доказом існування, поряд з вищеперерахованими типами семантичних структур, ще і комунікативно-прагматичних текстоцентричних структур. Текстоцентричне структурування наукової лексики проявляється в тому, що під час написання наукового тексту автор не лише вибирає терміни з відповідних семантико-темантичних груп, але, в залежності від того, для якого структурного компонента тексту вибираються терміни, він використовує дещо різну словникову базу.

При розгляді характерних особливостей творення наукового тексту, звертає на себе увагу той факт, що автор, присвячуючи текст певній науковій проблемі, подає її на більш широкому загальнонауковому тлі і, таким чином, прив'язує її до понять, які, на його думку, вже відомі потенційному читачеві. Починаючи текст, автор уникає вузькоспеціалізованих термінів і викладає проблему в більш загальновідомих словах. У наступних розділах, коли автор переходить до висвітлення специфічних для даного тексту питань, він поступово збагачує свій словник специфічними, щодо проблематики своєї статті, словами. У висновках, необхідність доведення до читача узагальненого змісту своєї роботи, змушує автора знову перейти до вживання більш загальних термінів. Умовно можна вважати, що під час написання різних розділів тексту /вступу, центральної частини та висновків/ автор користується різними словниками. У даній роботі проведено аналіз характерних семантичних особливостей цих словників і викладені результати дослідження взаємозалежності між семантичними та

дискурсивними характеристиками терміна, зокрема приналежністю терміну до тих чи інших компонентів лінійно-композиційної структури тексту.

Характерні особливості термінів та явища їх взаємодії з дискурсивними ознаками вивчалися методами кількісного аналізу. Перехід від описових семантичних характеристик термінів до їх кількісних еквівалентів представляє собою ряд формалізованих процедур з метою зробити процес кількісної оцінки однозначним і об'єктивним. У даному дослідженні використана модифікація схеми, яка успішно використовується для аналізу лексико-семантичних систем і відома під назвою "сіткове моделювання лексики" [Скороходько 1983: 218].

Ефективність застосування цієї схеми обумовлюється тим, що:

- семантична сітка є ідеальним засобом відображення будь-якого системно-структурного формування. Вона дозволяє в межах одного модельного об'єкту передати різноманітність мовної системи в повному обсязі як у цілому, так і в деталях з урахуванням характерних особливостей слів і їх зв'язків між собою, а також всіх інших лексичних структур більш високого рівня складності;
- відображення різнобічних, нерідко прихованих, семантичних зв'язків і семантичних структур термінології в явному вигляді, експліцитно, дає можливість перейти до об'єктивно кількісного опису лінгвістичних категорій, що, разом з застосуванням математичних, зокрема статистичних, методів під час аналізу лінгвістичних явищ, дає можливість більш глибоко проникати в їхню суть.

Сіткове моделювання термінологічної лексики починається з розкладання значень термінів на їхні семантичні складові. Припустимо, що існує деяка кількість слів, лексичні значення яких описані через семантичні складові або семантичні множини. Якщо лексичне значення одного слова безпосередньо або опосередковано є семантичною складовою другого (це виявляється зокрема у тому, що друге слово тлумачиться через перше), то це друге слово вважається семантичною похідною від першого (або його семантичним дериватом), а перше - семантичним компонентом другого. Цей етап роботи виконувався з використанням тлумачних словників. Вважається, що значення слів, які залучаються до тлумачення конкретного терміну, є його семантичними складовими [Скороходько 1968: 220]. Наприклад, семантичними складовими терміну САСНЕ є семантичні одиниці *memory*, *computer*, *instruction*, *program*, *code*, *area*, *block*. Таким чином, слово САСНЕ є семантичним дериватом слів *memory*, *computer*, *instruction*, *program*, *code*, *area*, *block*, а слова *memory*, *computer*, *instruction*, *program*, *code*, *area*, *block* є семантичними компонентами слова САСНЕ.

Наступний етап - це відображення робочого списку в рамках єдиної структури, яку називають "графом". Як математичний об'єкт [Оре 1965: 174], граф складається з вершин та ліній, що їх з'єднують. Ці лінії називаються ребрами. У даній роботі в якості вершин виступають терміни робочого списку, а ребра відображають наявність семантичного зв'язку між ними.

На малюнку 1 зображено фрагмент графу, до складу якого ввійшли 21 термін, тобто який складається з 21-ї вершини. Вершини графу з'єднують стрілки. Це і є

ребра графа. Термін, з якого виходить ребро, є семантичним компонентом терміну, в сторону якого це ребро направлено. І навпаки, термін, в сторону якого направлено дане ребро, є семантичним дериватом терміну, з якого це ребро виходить. Як показано на таблиці, totality є компонентом software, а software - є дериватом totality.

Наступний етап роботи полягає у підрахунку сумарного числа вершин, до яких можна дістатися рухаючись з даної вершини у напрямку протилежному напрямку ребер. Так підраховують кількість семантичних компонентів даного терміну. Сумарне число вершин, до яких можна дістатися рухаючись з даної вершини у напрямку ребер дає кількість семантичних дериватів даного терміну. Наприклад, в межах продемонстрованого фрагменту графу, кількість семантичних термінів: program дорівнює 4; command - 5; procedure - 1; operator - 6.



Мал.1. Фрагмент семантичної сітки.

В подальших розрахунках використовувалися як сумарна кількість семантичних компонентів, так і сумарна кількість семантичних дериватів тому, що цим параметрам легко дати лінгвістичну інтерпретацію [Скороходько 1997: 1-13]. Так, сумарну кількість семантичних компонентів можна трактувати як семантичну складність лексеми, а стосовно термінології, як інформативну місткість оскільки це кількість семантичних одиниць, які інтегруються у значенні даного терміну.

Сумарна кількість семантичних дериватів визначає кількість слів, у тлумаченні яких безпосередньо, або опосередковано використано даний термін і ця величина характеризує пояснювальну силу відповідної лексеми, або її семантичну продуктивність. Стосовно термінології, цю величину можна інтерпритувати, також, як когнітивну вартість терміну, оскільки вона визначає кількість понять, оволодіти якими неможливо, не засвоївши дане поняття.

В подальшому виконуються підрахунки для всього масиву вершин (всього 660 термінів). Результати розрахунків зводяться в таблицю, в якій кожному терміну робочого списку ставиться у відповідність сумарна кількість його семантичних компонентів та семантичних дериватів.

Останній етап роботи полягає в поширенні методики сіткового моделювання на лексику тексту. Семантичною складністю і когнітивною вартістю фрагменту

тексту в даному дослідженні є середня величина, яка дорівнює сумі семантичних складностей і когнітивних вартостей усіх термінів, що входять до цього фрагменту. Оцінка семантичних параметрів термінів в структурах тексту через середні величини накладає деякі обмеження як при обробці, так і при інтерпретації результатів. По-перше, процедури порівняння середніх величин повинні проводитися з урахуванням розподілу термінів в структурі тексту за семантичними параметрами. По-друге, отриманні результати підтверджують тільки ймовірний характер існування зв'язку між мовними та дискурсивними характеристиками терміну і дозволяють обчислити його достовірність.

Дослідження було проведено на масиві з 660 термінів, що були використані 3500 разів різними авторами при написанні трьох статей теоретичного спрямування з обчислювальної техніки. Терміни, що були використані у вступі всіх проаналізованих статей (умовно названі словником вступу), розглядалися як єдина вибірка. Із термінів центральної частини формувалася ще одна вибірка (словник центру). І, терміни, використані у висновках, склали ще одну окрему вибірку. Загальна множина фахових термінів, використаних в проаналізованих текстах, утворювала генеральний словник. Загальна характеристика когнітивної вартості та інформативної місткості термінів з обчислювальної техніки, що складають спільний для всіх текстів генеральний словник та словники вступу, центру та висновків, наведена у таблиці 1. Як видно з таблиці, лексичні обсяги словника центральних розділів та генерального словника приблизно рівні. З іншого боку зрозуміло, що за своїм лексичним складом словники вступу та висновків не можуть, у значній мірі, відрізнитися від словника центральних розділів. І дійсно, з 152 термінів, що складають

Таблиця 1

Загальні кількісні характеристики термінів обчислювальної та когнітивної вартості



інформативної місткості

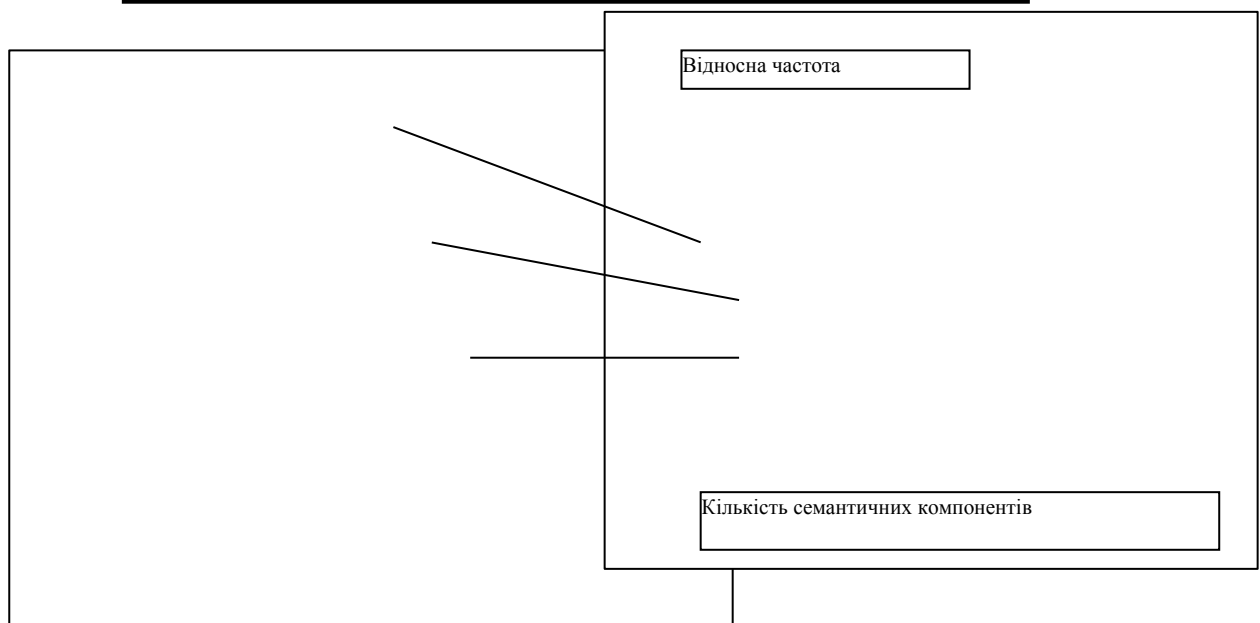


словник вступу, 82 входять до словника центру і лише 70 зовсім не зустрічається в ньому. Лінгвістичний аналіз об'єднаних словників показує, що під час написання центральних розділів практично повністю використовується термінологічна база генерального словника, тоді як при написанні словників вступу та висновків деякі терміни з генерального словника зовсім не використовуються. Перевірка за *t*-критерієм Стюдента [Тюрин,Макаров 1995: 187] показує, що різниці, які спостерігаються між словниками, не є випадковими (табл.2). Когнітивна вартість термінів словників вступу та висновків значно більша, ніж термінів словника центру. У той же час різниці в інформативних місткостях між словниками вступу та висновків мають меншу ймовірність, а в когнітивних вартостях взагалі відсутні.

Таблиця 2

Різниця між когнітивною вартістю та інформативною місткістю термінів словників вступу, центру та висновків

Порівнювані словники		Когнітивна вартість		Інформативна місткість	
		Різниця	Довірча ймовірність	Різниця	Довірча ймовірність
Центру	Вступу	-14,75	99,8%	2,77	99,9%
	Висновків	-22,05	99,9%	4,85	99,9%
Вступу	Висновків	-7,30	47,1%	2,09	80,7%



Мал.2 Функції розподілу термінів у вступі, висновках та центральних розділах текстів

Всі ці прояви текстоцентричного позиційного структурування наукової лексики найбільш яскраво спостерігається в функціях розподілу термінів за когнітивною вартістю та інформативною місткістю в відповідних словниках. На мал.2 функції

розподілу показано у вигляді графіків. Словник центру містить відносно більшу кількість термінів з малою когнітивною вартістю і відносно меншу кількість термінів з малою інформативною місткістю. Наприклад, терміни, що зовсім не мають семантичних дериватів, складають 76,4% словника центру і лише 56,6% та 44,3% словників вступу та висновків. Терміни з кількістю семантичних дериватів від 0 до 20 складають 95,0% словника центру і 86,2% та 83,6% словників вступу та висновків. Терміни з нульовою інформативною місткістю, навпаки, складають більший відсоток словників вступу та висновків ніж словника центру (відносно, 3,3%; 4,9%; 1,2%). Терміни, у яких кількість семантичних компонентів менша 5, складають 15,4% словника центру, 20,4% - словника вступу і 26,2% - словника висновків. Із співвідношень, які спостерігаються у функціях розподілу, можна зробити висновок, що:

- різниця середніх значень когнітивної вартості термінів у різних словниках обумовлюється, головним чином, різним розподілом, тобто, різним співвідношенням між кількістю термінів з великою когнітивною вартістю та кількістю термінів з малою когнітивною вартістю. Уявні лінії, що відображають функції розподілу, мають різну крутизну. Звідси і різниця у середніх значеннях: у словник центрального розділу потрапляє більше термінів з малою когнітивною вартістю, ніж у словники вступу чи висновків;

- терміни з малими значеннями когнітивної вартості відбираються переважно до словника центрального розділу, а терміни з малими значеннями інформативної місткості - переважно до словників вступу та висновків;

- терміни з великими значеннями когнітивної вартості, навпаки, відбираються переважно до словників вступу та висновків; а терміни з великими значеннями інформативної місткості - переважно до словника центрального розділу.

Таким чином, відбираючи терміни для написання різних розділів наукового тексту теоретичної спрямованості, автор, немовби, розчленовує словник (фактичний чи уявний), яким він користується, на окремі частини. Одну частину він використовує для написання вступу, другу - для центральних розділів і третю - для висновків. Ці частини можна умовно ідентифікувати як окремі словники - складові частини спільного (генерального) словника. Отримані, таким чином, словники вступу, центрального розділу та висновків не є жорстко розмежованими. Практично, словник центрального розділу містить у собі більшість термінів із словників вступу та висновків. У той же час, на відбір термінів до дискурсоорієнтованих словників впливають специфічні комунікативно-прагматичні установки, які обумовлюються різною комунікативною спрямованістю відповідних розділів тексту (ознайомлення читача з предметом обговорення, викладення основного змісту наукового дослідження, узагальнення). Тому у словники відбираються цілеспрямовано чи інтуїтивно дещо різні, за своїми системообумовленими характеристиками, терміни. Вивчення мови як системи або сукупності систем допомагає більш глибоко проникнути в сутність самої мови, висвітлити її внутрішню природу, розкрити первинні залежності між її

компонентами і механізми їх взаємодії в процесі творення цілісного тексту. Схема досліджень, що викладена в доповідді, є лише одним з варіантів реалізації даного підходу. Ця реалізація у варіанті сіткового моделювання при дослідженні лексики в цілому, а також особливостей її використання в текстах має певні переваги:

- семантичні зв'язки між словами відображаються у явному вигляді, що дає можливість формалізувати ряд процедур по аналізу семантичних зв'язків між словами та словосполученнями на рівні мови;

- схема сіткового моделювання дозволяє описувати систему семантичних структур лексики та тексту в системі кількісних сіткових параметрів, що, в свою чергу, дає можливість об'єктивно співставляти лексику як різних підмов, так і різних мов. Ця можливість особливо актуальна при цільовій спрямованості на лексику, наприклад, при впорядкуванні, або введенні в користування термінологічної лексики різних підмов;

- можливість визначення на кількісному рівні об'єктивних характеристик семантичних властивостей окремого слова в лексико-семантичній системі мови і в текстах разом з його частотними і змістовими характеристиками дозволяє спростити і формалізувати відбір лексики в словники різного призначення: машинні, які використовуються при автоматичній обробці текстів та термінологічні, тлумачні, перекладацькі.

І, нарешті, можливість визначення семантичного зв'язку між будь-якими одиницями лексико-семантичної системи та тексту дозволяє досліджувати взаємодію будь-якого елемента лексики як з будь-яким іншим елементом, так і з лексичною системою в цілому, що, в свою чергу, відкриває доступ до найбільш тонких механізмів функціонування окремих слів, або групи слів у комунікативних системах наукових текстів.

The interdependence between semantic complexity and cognitive value of the term and its position in the linear compositional structure of the text is discussed

Література

Оре О. Графы и их применение. - М.: Мир, 1965. - С. 174.

Скопюк Т.Г. Семантическая сложность термина и его позиция в научно-техническом тексте // Отраслевая терминология: Лингвопрагматические аспекты. Материалы межвузовской научно-практической конф. - Воронеж, 1997. - С. 40-45.

Скороходько Э.Ф. Семантические сети и автоматическая обработка текста. - Киев, 1983. - 218 с.

Скороходько Э.Ф., Пшеничная Л.Э., Медведев А.А. и др. Информационно-поисковая система. БИТ.. - Киев: Наукова думка, 1968. - С.220.

Скороходько Е.Ф. Семантическая продуктивность и семантическая емкость слова в общепотребительной и терминологической лексике: некоторые

количественные закономерности. - М.: Научно-техническая информация. Сер.2. Информационные процессы. - 1997. N. 2. - С.1-13.

Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. - М.: Финансы и статистика. - 1995. - С. 187.